15.

4. Find the minimizer of $c^T x$ (for $c \in \mathbb{R}^n$ a constant vector and $x \in \mathbb{R}^n$ a variable) over Ω, where Ω is each of the following sets:

(a) The unit ball: $\{x \mid \|x\|_2 \leq 1\}$;
(b) The unit simplex: $\{x \in \mathbb{R}^n \mid x \geq 0, \ \sum_{i=1}^n x_i = 1\}$;
(c) A box: $\{x \mid 0 \leq x_i \leq 1, \ i = 1, 2, \dots, n\}$.

1. $X^* = \underset{X \in \Omega}{\operatorname{argmin}} \ C^T X$ $\qquad \phi(X) = C^T X$ is also a convex function

$\nabla f(X^*) = C$ $\qquad$ we have $\ -\nabla f(X^*) \in N_\Omega(X^*)$

$\forall u \in \Omega: \quad \langle C, u - X^* \rangle \geq 0$

(a) for unit ball $\{X : \|X\|_2 \leq 1\}$

then we have $\quad X^* = -\dfrac{C}{\|C\|_2}$

$C^T X^* = -\|C\|_2$

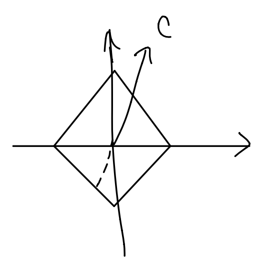
$\forall u \in \Omega. \ |C^T u| \leq \|C\|_2 \|u\|_2 \leq \|C\|_2$ $\qquad$ 5.

$C^T u \geq -\|C\|_2 = C^T X^* \Rightarrow \langle C, u - X^* \rangle \geq 0$

(b) $\Omega = \{X \in \mathbb{R}^n \mid X \geq 0, \ \sum_{i=1}^n X_i = 1\}$

$X^* = e_{\underset{1 \leq i \leq n}{\operatorname{argmin}} C_i}$

$X_i^* = \begin{cases} 1 & i = \underset{j}{\operatorname{argmin}} \ C_j \quad j = 1 \cdots n \\ 0 & i \neq \underset{j}{\operatorname{argmin}} \ C_j \quad j = 1 \cdots n \end{cases}$



for $\forall u \in \Omega$. suppose $C_i = \min_{1 \leq j \leq n} C_j$. then: $X_i^* = 1 \quad X_j = 0 \ (j \neq i)$

$u - X^* = (u_1 \cdots, u_{i-1}, u_{i+1}, \cdots u_n)$ $\qquad$ 5.

$\langle C, u - X^* \rangle = \sum_{j=1}^n C_j u_j - C_i \geq \sum_{j=1}^n C_i u_j - C_i = 0$

(c) $\Omega = \{X \mid 0 \leq X_i \leq 1, \ i = 1 \cdots n\}$

$X_i^* = \begin{cases} 0 & C_i \geq 0 \quad i = 1 \cdots n \\ 1 & C_i \leq 0 \quad i = 1 \cdots n \end{cases}$

$\forall u \in \Omega.$ $\quad u_i \in [0,1]$ $\quad u_i - 1 \le 0$

$$\langle c, u - x^* \rangle = \sum_{i: c_i \ge 0} c_i u_i + \sum_{j: c_j \le 0} c_i (u_i - 1)$$

≥ 0

5.

10.

2. **Q 2.** Consider the unconstrained optimization problem $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$, where f is an L-smooth convex function. Assume that $\|\mathbf{x}_0 - \mathbf{x}^*\|_2 \le R$, for some $R \in (0, \infty)$, and let $f_\epsilon(\mathbf{x}) = f(\mathbf{x}) + \frac{\epsilon}{2R^2}\|\mathbf{x} - \mathbf{x}_0\|_2^2$. Let $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$ and $\mathbf{x}_\epsilon^* = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} f_\epsilon(\mathbf{x})$. You have already shown in previous homework (with possibly minor modifications) that:

$$(\forall \mathbf{x} \in \mathbb{R}^n) : f(\mathbf{x}) - f(\mathbf{x}^*) \le f_\epsilon(\mathbf{x}) - f_\epsilon(\mathbf{x}_\epsilon^*) + \frac{\epsilon}{2}.$$

(i) Prove that Nesterov's method for smooth and strongly convex minimization applied to f_ϵ will find a solution $\mathbf{x}_k$ with $f(\mathbf{x}_k) - f(\mathbf{x}^*) \le \epsilon$ in $O(\sqrt{\frac{L}{\epsilon}} R \log(\frac{LR^2}{\epsilon}))$ iterations. [5pts]

(ii) Using the lower bound for smooth minimization we have proved in class, prove the following lower bound for L-smooth and m-strongly convex optimization: any method satisfying the same assumption as we used in class (that $\mathbf{x}_k \in \mathbf{x}_0 + \mathrm{Lin}\{\nabla f(\mathbf{x}_0), \ldots, \nabla f(\mathbf{x}_{k-1})\}$) must take at least $k = \Omega(\sqrt{\frac{L}{m}})$ iterations in the worst case to construct a point $\mathbf{x}_k$ such that $f(\mathbf{x}_k) - f(\mathbf{x}^*) \le \epsilon$, for any $\epsilon > 0$. [10pts]

(i) From the lecture, we have:

Given a smooth strongly convex function: apply (A+AGD) we have:

$f(x_k^{out}) - f(x^*) \le \varepsilon$ in

$$O\left(\sqrt{\frac{L}{m}} \log_2\left(\frac{L\|x_0 - x^*\|_2^2}{\varepsilon}\right)\right) \text{ iterations}$$

where f is L-smooth and m-convexity

Since $f_\varepsilon(x) = f(x) + \frac{\varepsilon}{2R^2}\|x - x_0\|_2^2$

$\nabla f_\varepsilon(x) = \nabla f(x) + \frac{\varepsilon}{R^2}(x - x_0)$

$\forall x, y \in R^n$

$$\|\nabla f_\varepsilon(y) - \nabla f_\varepsilon(x)\| \le \|\nabla f(y) - \nabla f(x)\| + \frac{\varepsilon}{R^2}\|y - x\|$$

$$\le \left(L + \frac{\varepsilon}{R^2}\right)\|y - x\| \quad \Rightarrow \quad L^\varepsilon = L + \frac{\varepsilon}{R^2}$$

$$f_\varepsilon(y) - f_\varepsilon(x) = f(y) - f(x) + \frac{\varepsilon}{2R^2}\|y - x_0\|_2^2 - \frac{\varepsilon}{2R^2}\|x - x_0\|_2^2$$

According to Taylor's theorem:

$$f_\varepsilon(y) - f_\varepsilon(x) \geq \langle \nabla f(x), y-x \rangle + \langle \frac{\varepsilon}{R^2}(x-x_0), y-x \rangle + \frac{\varepsilon}{2R^2}\|y-x\|_2^2$$

$$= \langle \nabla f_\varepsilon(x), y-x \rangle + \frac{\varepsilon}{2R^2}\|y-x\|_2^2$$

$$\Rightarrow \quad m^\varepsilon = \frac{\varepsilon}{R^2}$$

Then applying (A+AGD) we have:

$f_\varepsilon(x_k^{out}) - f_\varepsilon(x^*) \leq \frac{\varepsilon}{2}$ in

$$O\left(\sqrt{\frac{L+\frac{\varepsilon}{R^2}}{\frac{\varepsilon}{R^2}}} \log\left(\frac{2\left(L+\frac{\varepsilon}{R^2}\right)R^2}{\varepsilon}\right)\right)$$

$$= O\left(\sqrt{\frac{L}{\varepsilon}} R \log\left(\frac{LR^2}{\varepsilon}\right)\right) \quad \text{iterations}$$

also: $f(x_k^{out}) - f(x^*) \leq f_\varepsilon(x_k^{out}) - f_\varepsilon(x^*) + \frac{\varepsilon}{2} \leq \varepsilon$

5.

(ii) According to the theorem: for any $1 \leq k \leq \frac{1}{2}(n-1)$, exists a function $f \in F_L^{\infty,1}(\mathbb{R}^n)$

any method satisfy the assumption:

$$f(x_k) - f(x^*) \geq \frac{3L\|x_0-x^*\|_2^2}{32(k+1)^2} \geq \varepsilon_0$$

$$k+1 = \sqrt{\frac{3L\|x_0-x^*\|_2^2}{32\varepsilon_0}}$$

Consider $k = \frac{1}{100}\sqrt{\frac{L}{m}} - 1$

$$f(x_k) - f(x^*) \geq \frac{3m}{32} \cdot 10000\|x_0 - x^*\|_2^2$$

Not sufficient as a proof. See Solution!

$\exists\ \varepsilon = 3m\|x_0 - x^*\|_2^2$ s.t.

$f(x_k) - f(x^*) \leq \varepsilon$ must take $k = \Omega\left(\sqrt{\frac{L}{m}}\right)$

5.

Q 4. In this part, you will compare the heavy ball method to Nesterov's method for smooth and strongly convex optimization. Your problem instance is the following one-dimensional instance: $\min_{x\in\mathbb{R}} f(x)$, where

$$f(x) = \begin{cases} \frac{25}{2}x^2, & \text{if } x < 1 \\ \frac{1}{2}x^2 + 24x - 12, & \text{if } 1 \le x < 2 \\ \frac{25}{2}x^2 - 24x + 36, & \text{if } x \ge 2. \end{cases}$$

Prove that f is m-strongly convex and L-smooth with $m = 1$ and $L = 25$. What is the global minimizer of f? (Justify your answer.)

Run Nesterov's method and the heavy-ball method, starting from $x_0 = 3.3$. Plot the optimality gap of Nesterov's method and the heavy ball method over 100 iterations. What do you observe? What does this plot tell you? [30pts]

4. Since f is smooth and strongly convex:

Consider Taylor's theorem. for $\forall x, y \in R^1$ $\quad p = y - x$

$$|\nabla f(y) - \nabla f(x)| = \left|\int_0^1 \nabla^2 f(x + rp)\, p\, dr\right|$$

$$\le \left|\int_0^1 25\, p\, dr\right| \le 25|y-x|$$

can use Taylor only if f'' is continuous

-5

For $\forall x, y \in R^1$ $\quad p = y - x$

$$f(y) - f(x) - \nabla f(x) \cdot p \ge \frac{1}{2}\nabla^2 f(x) \cdot p^2$$

$$\ge \frac{1}{2}p^2$$

Then f is m- convexity and L smooth.

$L = 25$. $\quad m = 1$

$$\nabla f(x) = \begin{cases} 25x & x < 1 \\ x + 24 & 1 \le x < 2 \\ 25x - 24 & 2 \le x \end{cases}$$

$\nabla f(x) = 0 \Rightarrow x^* = 0 \quad \nabla f(x^*) = 0$

f is strongly convex.

$$\lim_{\substack{x \to +\infty \\ x \to -\infty}} f(x) = +\infty \qquad f(0) = 0$$

Then the global minimizer is $x^* = 0$ is unique

Please see jupyter for plots.

15.

Q 5. Suppose that I give you an algorithm (let's call it AGD-G) that given an initial point $\mathbf{x}_0 \in \mathbb{R}^n$ and gradient access to an L-smooth function $f : \mathbb{R}^n \to \mathbb{R}$ (where $0 < L < \infty$) after k iterations returns a point $\mathbf{x}_k \in \mathbb{R}^n$ that satisfies:

$$\|\nabla f(\mathbf{x}_k)\|_2 \leq \sqrt{\frac{2L(f(\mathbf{x}_0) - f(\mathbf{x}^*))}{(k+1)^2}}.$$

Note that AGD-G does not need to know the value of L.

Show that you can use AGD-G to obtain an algorithm that for any m-strongly convex and L-smooth function and any $\epsilon > 0$ can construct a point $\mathbf{x}_k$ with $f(\mathbf{x}_k) - f(\mathbf{x}^*) \leq \epsilon$ in $k = O\Big(\sqrt{\frac{L}{m}} \log\big(\frac{L\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{\epsilon}\big)\Big)$ iterations. Your algorithm should work without the knowledge of the values of L and m. [15pts]

Extra: $f(x_0) - f(x^*) \leq \frac{L}{2}\|x_0 - x^*\|_2^2$

Since (AGD-G) have:

$$\|\nabla f(x_k)\|_2 \leq \sqrt{\frac{2L(f(x_0) - f(x^*))}{(k+1)^2}} \leq \frac{L\|x_0 - x^*\|}{k+1}$$

$$f(x^*) - f(x_k) \geq \langle \nabla f(x_k), x^* - x_k \rangle$$

$$f(x_k) - f(x^*) \leq \langle \nabla f(x_k), x_k - x^* \rangle \leq \|\nabla f(x_k)\|_2 \cdot \|x_k - x^*\|_2$$

$$\leq \frac{L}{k+1}\|x_0 - x^*\|_2^2$$

$$f(x^*) \geq f(x_k) + \langle \nabla f(x_k), x^* - x_k \rangle + \frac{m}{2}\|x^* - x_k\|_2^2$$

$$f(x_k) - f(x^*) \leq \langle \nabla f(x_k), x_k - x^* \rangle - \frac{m}{2}\|x_k - x^*\|_2^2$$

$$\leq -\frac{m}{2}\|x_k - x^*\|_2^2 + \|\nabla f(x_k)\| \cdot \|x_k - x^*\|$$

$\downarrow$ $p = \sqrt{m}\|x_k - x^*\|$ $\quad q = \frac{1}{\sqrt{m}} \cdot \|\nabla f(x_k)\|$

$$\leq \frac{1}{2m}\|\nabla f(x_k)\|_2^2$$

$$\leq \frac{L^2}{2m(k+1)^2}\|x_0 - x^*\|_2^2$$

$$f(x_k) \geq f(x^*) + \frac{m}{2}\|x_k - x^*\|_2^2$$

$$\frac{m}{2}\|x_k - x^*\|_2^2 \leq f(x_k) - f(x^*) \leq \frac{L^2}{2m(k+1)^2}\|x_0 - x^*\|_2^2$$

$$\|x_k - x^*\|_2^2 \leq \left(\frac{L}{m}\right)^2 \cdot \frac{1}{(k+1)^2}\|x_0 - x^*\|_2^2$$

when $k \geq \frac{2L}{m}$. we have: $\|x_k - x^*\|_2^2 \leq \frac{1}{2}\|x_0 - x^*\|_2^2$

B: for $j = 1 : G$

$$x_j^{ont} = AGD - G\left(x_{j-1}^{ont},\ k = \frac{2L}{m}\right)$$

Do you know $\frac{L}{m}$?

WANT: $\|x_j^{ont} - x^*\|_2^2 \le \varepsilon$

$$G \geq \log_2\left(\frac{\|x_0 - x^*\|_2^2}{\varepsilon}\right)$$

$$\left(\frac{1}{2}\right)^G \|x_0 - x^*\|_2^2 \le \varepsilon$$

X

then:

$$f(x_G^{ont}) - f(x^*) \le \frac{L}{2}\|x_G^{ont} - x^*\|_2^2 \le \frac{L}{2}\varepsilon = \bar{\varepsilon}$$

Let $\bar{\varepsilon} = \frac{L}{2}\varepsilon$ $\qquad \varepsilon = \frac{2}{L}\bar{\varepsilon}$

Then: $f(x_G^{ont} - x^*) \le \bar{\varepsilon}$

when $G \geq \log_2\left(\frac{L\|x_0 - x^*\|_2^2}{2\bar{\varepsilon}}\right)$

Total At AGD-G:

$$K = O\left(\frac{L}{m} \cdot \log_2 \frac{L\|x_0 - x^*\|_2^2}{\bar{\varepsilon}}\right)$$

8.